nature genetics

A variant of the gene encoding leukotriene A4 hydrolase confers ethnicity-specific risk of myocardial infarction

Anna Helgadottir[1], Andrei Manolescu[1], Agnar Helgason[1], Gudmar Thorleifsson[1], Unnur Thorsteinsdottir[1], Daniel F Gudbjartsson[1], Solveig Gretarsdottir[1], Kristinn P Magnusson[1], Gudmundur Gudmundsson[1], Andrew Hicks[1], Thorlakur Jonsson[1], Struan F A Grant[1], Jesus Sainz[1], Stephen J O'Brien[2], Sigurlaug Sveinbjornsdottir[3], Einar M Valdimarsson[3], Stefan E Matthiasson[3], Allan I Levey[4], Jerome L Abramson[4], Murdach P Reilly[5], Viola Vaccarino[4], Megan L Wolfe[5], Vilmundur Gudnason[6], Arshed A Quyyumi[4], Eric J Topol[7], Daniel J Rader[5], Gudmundur Thorgeirsson[3], Jeffrey R Gulcher[1], Hakon Hakonarson[1], Augustine Kong[1] & Kari Stefansson[1]

Variants of the gene *ALOX5AP* (also known as *FLAP*) encoding arachidonate 5-lipoxygenase activating protein are known to be associated with risk of myocardial infarction[1]. Here we show that a haplotype (HapK) spanning the *LTA4H* gene encoding leukotriene A4 hydrolase, a protein in the same biochemical pathway as *ALOX5AP*, confers modest risk of myocardial infarction in an Icelandic cohort. Measurements of leukotriene B4 (LTB4) production suggest that this risk is mediated through upregulation of the leukotriene pathway. Three cohorts from the United States also show that HapK confers a modest relative risk (1.16) in European Americans, but it confers a threefold larger risk in African Americans. About 27% of the European American controls carried at least one copy of HapK, as compared with only 6% of African American controls. Our analyses indicate that HapK is very rare in Africa and that its occurrence in African Americans is due to European admixture. Interactions with other genetic or environmental risk factors that are more common in African Americans are likely to account for the greater relative risk conferred by HapK in this group.

To search for SNPs and potential causal variants of *LTA4H*, we sequenced DNA across the *LTA4H* gene region (42 kb) in 93 individuals affected with myocardial infarction. Although no coding sequence variant leading to amino acid substitutions was found, we identified and selected eight SNPs and genotyped them, together with two known SNPs in the 5′ region of the gene (**Fig. 1**), in Icelandic individuals with myocardial infarction and controls. These SNPs extend 11.9 kb upstream and 1 kb downstream of the *LTA4H* coding sequence and were selected to capture all haplotypes with a frequency of >2% across the gene region.

We tested the ten SNPs for association with myocardial infarction by using 1,553 individuals with myocardial infarction and 863 population-based controls. No single SNP or haplotype defined by the ten SNPs was found to be significantly more common in all individuals with myocardial infarction than in controls (**Supplementary Tables 1** and **2** online). Therefore, we tested association of the haplotypes with more severe myocardial infarction phenotypes—namely, early-onset myocardial infarction and myocardial infarction with other cardiovascular diseases, including peripheral vascular disease, stroke, or both. Early-onset myocardial infarction did not show significant association with any of the haplotypes (data not shown); however, myocardial infarction with additional cardiovascular diseases showed association with a haplotype that we called HapK (**Fig. 1** and **Table 1**). The frequency of HapK in individuals with myocardial infarction and additional cardiovascular disease and in controls was 14.5% and 10.4%, respectively, corresponding to a relative risk of 1.45 ($P = 0.0091$) for each copy of HapK carried ($P = 0.035$ after adjusting for the number of haplotypes tested).

To investigate the functional relevance of HapK, we examined the correlation between HapK carrier status and the amount of LTB4, the main product of the LTA4H enzyme, that was produced by granulocytes isolated from the same individuals. We have previously reported[1] that granulocytes from individuals with myocardial infarction ($n = 41$) produce more LTB4 than those from controls without any history of myocardial infarction ($n = 36$). This data set included 14 HapK carriers: seven individuals with myocardial infarction (one homozygote) and seven controls. Using multiple regression including age, gender and disease status as covariates, we observed a positive correlation between HapK and LTB4 production after stimulating the cells for 15 min ($P = 0.015$) and 30 min ($P = 0.009$) with ionomycin (**Table 3** and **Supplementary Table 3** online).

[1]deCODE Genetics, Inc., Sturlugata 8, IS-101Reykjavik, Iceland. [2]Laboratory of Genomic Diversity, National Cancer Institute, Frederick, Maryland 21702, USA. [3]National University Hospital, Reykjavik, Iceland. [4]Emory University School of Medicine, Atlanta, Georgia, USA. [5]University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA. [6]Icelandic Heart Association, Holtasmári 1, 201 Kópavogur, Reykjavik, Iceland. [7]Cleveland Clinic Foundation, 9500 Euclid Avenue, Cleveland, Ohio 44195, USA. Correspondence should be addressed to K.S. (kstefans@decode.is).

Received 8 June; accepted 5 October; published online 10 November 2005; doi:10.1038/ng1692

Given the modest risk conferred by HapK in Iceland, we performed a replication study in three independent myocardial infarction cohorts from the United States recruited in Philadelphia, Cleveland and Atlanta. All three cohorts contained both self-reported European Americans and African Americans (**Table 1**), who were analyzed separately. **Table 1** shows the association results for HapK in each of these cohorts. The *P* values reported for all of the replication analyses are one sided because we tested only HapK for increased risk. An excess of HapK was detected in European American individuals with myocardial infarction from Philadelphia (relative risk = 1.37, $P = 0.0051$) and Cleveland (relative risk = 1.12, not significant), but not in those from Atlanta (**Table 1**). The association of HapK with myocardial infarction in European Americans was replicated when the three cohorts were simply combined (relative risk = 1.19, $P = 0.006$), and when a Mantel-Haenszel–like[2] analysis was done to adjust for differences in HapK frequency among controls in the three cohorts (relative risk = 1.16, $P = 0.019$; **Table 2**). As in Iceland, the risk of HapK was greater in those individuals with myocardial infarction who had a history of stroke or peripheral vascular disease (**Table 1**), with the combined cohort adjusted analysis yielding a relative risk of 1.31 ($P = 0.037$; **Table 2**).

Although HapK was found to be less frequent in African Americans (**Table 1**), its association with myocardial infarction was much stronger in this group, with the relative risk estimated as 6.50, 1.78 and 5.21 for the cohorts from Philadelphia, Cleveland and Atlanta, respectively (**Table 1**). The estimated relative risk was substantially less in Cleveland than in the other two cohorts, mainly because the control frequency of HapK is greater in that cohort. The relative risk conferred by HapK in the combined group of all African Americans with cohort adjustment was estimated to be 3.57 ($P = 0.000022$). Its confidence interval did not overlap with that of the European Americans (**Table 2**), showing that the relative risk of HapK in these two groups is significantly different ($P < 0.001$).

As HapK is much more frequent in European Americans than in African Americans, it is possible that the greater relative risk of myocardial infarction in African Americans is in part attributable to a greater European ancestry in individuals with myocardial infarction than in controls. This could be caused either by a bias in data collection (such as differences in recruitment of the myocardial infarction and control groups), or because European ancestry itself is a risk factor for myocardial infarction in African Americans or a close correlate of such a risk factor. To investigate this further, we genotyped a set of 75 unlinked microsatellite markers, selected as informative for distinguishing between African and European ancestry (see Methods and **Supplementary Table 4** online), in the three US cohorts, in 364 Icelanders and in 90 Nigerian Yorubans used in the HapMap project[3]. We used Structure software[4,5] to analyze these data to estimate the distribution of European ancestry in individuals grouped by disease status and self-reported ethnicity (**Table 4**). We also obtained estimates of European ancestry by applying a weighted least-squares (WLS) estimator[6] to a subset of the microsatellite alleles that showed the greatest differences in frequency between European and African populations in accordance with ref. 7 (**Table 4**). Overall, we found a close correspondence between self-reported ethnicity and the estimated ancestry derived from the genetic markers and also between the estimated individual ancestry (Structure) and group ancestry (WLS). In particular, the almost perfect assignment of African ancestry to Nigerian Yorubans and European ancestry to Icelanders indicated that the admixture estimates of the American individuals with myocardial infarction and controls were reliable. Furthermore, our estimates of European ancestry in African Americans were in the range reported in most previous studies[7–11].

Notably, we found that African American individuals with myocardial infarction had, on average, a slightly greater European ancestry than did the African American controls in the Philadelphia and Atlanta cohorts (**Table 4**). When all three cohorts were combined, the African American individuals with myocardial infarction and controls had on average 22.3% and 19.9% European ancestry, respectively (one-sided $P = 0.046$). This difference can largely be accounted for by a few individuals who were recorded as African Americans but had a relatively large European ancestry. We corrected for potentially misclassified individuals by excluding from the study self-reported African Americans with <20% African genetic ancestry according to the Structure results (seven individuals with myocardial infarction and four controls). The result was a notable reduction in the difference between individuals with myocardial infarction (20%) and controls (19.2%). Controlling for ancestry, whether by excluding potentially misclassified individuals or by using individual European ancestry estimates as covariate[12], referred to as 'admixture adjustment', has a negligible effect on the relative risk and statistical significance of the association of HapK with myocardial infarction in African Americans (**Tables 1** and **2**). We conclude that the higher relative risk of HapK in African Americans is not simply a consequence of differences in European ancestry between individuals with myocardial infarction and controls.

Notably, however, African American carriers of HapK had, on average, more European ancestry than those who did not carry HapK: 28.9% versus 19.8% (two-sided $P = 0.00008$). This is consistent with the observation that HapK was not found in the Nigerian HapMap sample, but was relatively common in the Icelandic and the CEPH CEU (Utah residents with ancestry from northern and western Europe) samples used in the HapMap project (**Supplementary Fig. 1** and **Table 2** online). Although HapK was found to be common in the Asian HapMap samples, the Structure-based estimate of Asian ancestry in African Americans was small (~1%), supporting the hypothesis that copies of HapK present in African Americans are mostly of European origin. Furthermore, we detected no difference in Asian ancestry between African American individuals with myocardial infarction and controls or between HapK carriers and noncarriers.

The *LTA4H* gene is located in a single linkage disequilibrium (LD) block in both European and African populations and is the only gene known in that block (**Supplementary Fig. 2** online). To identify a single causal variant captured by HapK, we sequenced a region of 75 kb

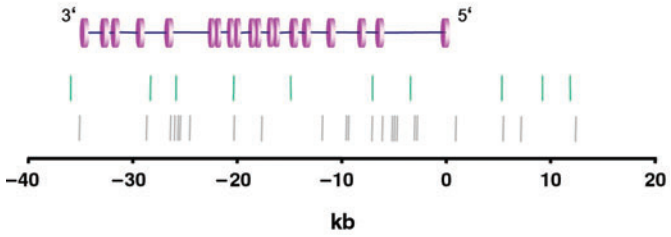

Figure 1 Structure of the *LTA4H* gene. Exons are shown as pink cylinders, and the positions of all genotyped SNPs relative to exons are shown as green lines. The SNPs and alleles (defined on the plus strand) defining HapK are SG12S16 (C), rs2660880 (G), rs6538697 (T), rs1978331 (A), rs17677715 (T), rs2247570 (T), rs2660898 (T), rs2540482 (C), rs2660845 (G) and rs2540475 (G). See information on SG12S16 in **Supplementary Table 1**. The relative position of SNPs typed in the HapMap project[3] (Phase I, version 16c.1) are shown as gray lines. For Icelanders and European Americans, the association results in **Tables 1** and **2** could be obtained with only five SNPs (rs1978331, rs17677715, rs2540482, rs2660845 and rs2540475). For African Americans, because of admixture effects, two more SNPs (rs2247570, rs2660898) had to be added to the above five to reproduce the results obtained with HapK.

encompassing the LD block containing *LTA4H* in several pooled DNA samples of Icelandic individuals with myocardial infarction and controls. Some pooled samples contained only HapK carriers. In addition, we examined the correlation of HapK with other SNPs in the HapMap[3] database (Phase I, version 16c.1). The best single SNP surrogate of HapK identified through both of these approaches was rs2660899 ($R^2 = 0.7$ in the CEU samples). We genotyped this SNP in the Philadelphia cohort, in which HapK showed the strongest effect. Although the T allele conferred a relative risk of 1.31 ($P = 0.008$) in European Americans, it did not fully capture the disease association with HapK in this African American

Table 1 Association of HapK with myocardial infarction

	Frequency of HapK			
Cohorts (*n*)	Individuals with MI	Controls	Relative risk	*P* value[a]
Icelanders				
All MI (1,553/863)	0.113	0.104	1.1	0.36
MI and additional CVD (325/863)	0.145	0.104	1.45	0.0091
European Americans				
Philadelphia				
All MI sre (728/430)	0.186	0.143	1.37	0.0051
All MI gda (724/430)	0.186	0.143	1.37	0.0051
All MI admix adj			1.36	0.0048
Cleveland				
All MI sre (627/792)	0.166	0.151	1.12	0.15
All MI gda (626/792)	0.166	0.151	1.11	0.16
All MI admix adj			1.12	0.15
MI and additional CVD sre (144/792)	0.193	0.151	1.34	0.046
MI and additional CVD admix adj			1.34	0.044
Atlanta				
All MI sre (236/553)	0.135	0.143	0.94	0.64
All MI gda (236/553)	0.135	0.143	0.94	0.64
All MI admix adj			0.94	0.63
MI and additional CVD sre (39/553)	0.173	0.143	1.25	0.25
MI and additional CVD admix adj			1.24	0.26
African Americans				
Philadelphia				
All MI sre (105/127)	0.103	0.017	6.5	0.000067
All MI gda (100/126)	0.104	0.018	6.45	0.000088
All MI admix adj			6.34	0.00010
Cleveland				
All MI sre (53/111)	0.122	0.072	1.78	0.11
All MI gda (52/111)	0.112	0.072	1.61	0.17
All MI admix adj			1.75	0.11
MI and additional CVD sre (13/111)	0.152	0.072	2.31	0.14
MI and additional CVD admix adj			2.27	0.16
Atlanta				
All MI sre (39/149)	0.075	0.015	5.21	0.018
All MI gda (38/146)	0.071	0.016	4.71	0.025
All MI admix adj			5.08	0.019
MI and additional CVD sre (8/149)	0.202	0.015	16.36	0.0039
MI and additional CVD admix adj			16.67	0.0035

Shown is the frequency of HapK in individuals with myocardial infarction (MI) and controls, together with the corresponding numbers (*n*) of subjects (individuals with myocardial infarction/controls), the relative risk and *P* values. Results are shown for European Americans and African Americans, defined by their self-reported ethnicity (sre). For each self-reported group, results are also shown for those who had a genetically detected ancestry (gda) of at least 20% European (in European Americans) and at least 20% African (in African Americans). Results adjusted for admixed ancestry in each self-reported group are also shown (admix adj). Myocardial infarction and additional cardiovascular diseases (CVD) refer to those individuals with myocardial infarction who also had either peripheral vascular disease or who had suffered a stroke. Information on previous history of stroke or peripheral vascular disease was not available for the subjects from Philadelphia.

[a]*P* values are two-sided for Icelanders but one-sided in all the other cohorts because we specifically tested the excess of HapK in individuals with myocardial infarction relative to controls.

Table 2 Association of HapK with myocardial infarction in combined American cohorts

Ethnic groups (*n*)	Frequency of HapK		RR (95% CI)	*P* value	PAR
	Individuals with MI	Controls			
European Americans					
All MI (1,591/1,775)	0.171	0.148	1.19 (1.04, 1.36)	0.006	
All MI coh adj			1.16 (1.01, 1.34)	0.019	
All MI coh adj, admix adj			1.16 (1.01, 1.33)	0.017	0.046
MI and additional CVD (183/1345)[a]	0.192	0.15	1.35 (1.00, 1.81)	0.026	
MI and additional CVD coh adj			1.31 (0.97, 1.78)	0.037	
MI and additional CVD coh adj, admix adj			1.32 (0.98, 1.78)	0.035	0.089
African Americans					
All MI (197/387)	0.105	0.032	3.52 (1.96, 6.29)	0.000012	
All MI coh adj			3.57 (1.94, 6.57)	0.000022	
All MI coh adj, admix adj			3.50 (1.90, 6.43)	0.000029	0.144
MI and additional CVD (21/260)[a]	0.176	0.041	4.94 (1.58, 15.43)	0.003	
MI and additional CVD coh adj			4.39 (1.32, 14.64)	0.008	
MI and additional CVD coh adj, admix adj			4.17 (1.21, 14.30)	0.012	0.219

The results describe the association of HapK with myocardial infarction (MI) in combined groups of self-reported European and African Americans from Philadelphia, Cleveland and Atlanta. The haplotype frequencies, the relative risk (RR) and the *P* values are shown first without any population adjustment; second, after adjusting for different cohort or population frequencies (coh adj); and third, after further adjusting for the admixture of African and European ancestries in each ethnic group (admix adj). All *P* values are one sided. PAR is the population attributable risk. CI, confidence interval.

[a]Only the Cleveland and Atlanta cohorts were combined for the severe phenotype of myocardial infarction and additional cardiovascular disease, as this information was not available for the subjects from Philadelphia.

cohort (**Supplementary Fig. 3** online). Thus, rs2660899 can be ruled out as a sole causal variant captured by HapK.

In theory, the observed association of myocardial infarction with HapK could be the result of an association with a causal variant located in the neighborhood of *LTA4H* but outside the LD block. Such a situation might explain the high relative risk observed in the recently admixed African Americans, potentially boosted by strong admixture-derived LD, and the modest relative risk in the nonadmixed groups of European Americans and Icelanders. Given the existing patterns of LD in European and African populations, however, the kind of admixture found in African Americans, which we examined by creating a 4:1 mixture of haplotypes from the Yoruban and CEPH CEU HapMap samples, would not be expected to produce a correlation ($R^2 > 0.25$) between HapK and any known SNP outside the *LTA4H* LD block. Because the observed effect of HapK on myocardial infarction is very strong in African Americans, it is implausible that the association is the consequence of a variant that is only loosely correlated with HapK. In addition, in an analysis of five markers located just outside the *LTA4H* LD block with significant allele frequency differences between African and European American controls, none was associated with HapK or differed between African American individuals with myocardial infarction and controls (**Supplementary Table 5** online). Thus, the difference in ancestry between African American individuals with myocardial infarction and controls seems to be localized to HapK.

The identification of a genetic variant that confers such different risks of myocardial infarction in African Americans and populations of European descent suggests a strong interaction between HapK and other genetic variants and/or non-genetic risk factors that are more common in African Americans than in European Americans and Icelanders. Our results emphasize that although genetic differences between human continental groups are small[13,14], some of these differences may nonetheless contribute to ethnicity-based health disparities[15], whether through frequencies of risk alleles, through risk conferred by such alleles, or both. We and others[16] have found a strong correspondence between self-reported ethnicity and genetically estimated ancestry. However, ancestry is a quantifiable trait, particularly in heterogeneous or recently admixed populations such as African Americans, that needs to be assessed to interpret reliably interactions among ancestry, genes and environment in the pathogenesis of disease[11,17,18].

Several reports indicate that the leukotriene pathway has a role in the pathogenesis of atherosclerosis, in particular in the branch involved in LTB4 biosynthesis[19–21]. We have shown that HapK is correlated with risk of myocardial infarction and increased production of LTB4, the main product of the enzyme encoded by *LTA4H*. LTB4 produced through activation of the leukotriene pathway may amplify inflammatory responses in the arterial wall, by mediating chemotaxis and thereby promoting adhesion of leukocytes to the vascular endothelium and transmigration. In addition, LTB4-induced activation of leukocytes leads to the release of lysosomal enzymes such as myeloperoxidase and the generation of reactive oxygen species[22], which have been attributed to initiation, propagation and acute complications of atherosclerosis[23,24]. Overall, these findings suggest that agents affecting LTB4 biosynthetic pathways may prove useful for primary or secondary prevention of heart attacks.

METHODS

Subjects from Iceland. The study cohort comprised 1,553 unrelated Icelandic subjects with myocardial infarction, including 597 with early-onset myocardial infarction and 325 with additional atherosclerotic manifestations (stroke and/or peripheral arterial disease), and 863 unrelated population controls. Recruitment

Table 3 Correlation between LTB4 and myocardial infarction and HapK carrier status[a]

	After 15 min	After 30 min
Predictor variable[a]	*P* value	*P* value
Disease status[b]	0.011	0.016
Carriers of HapK	0.015	0.009

[a]Two-sided *P* values correspond to a correlation between LTB4 after ionomycin stimulation of isolated granulocytes and both myocardial infarction status and the carrier status of the at-risk haplotype HapK. The results for the two time points were calculated by multiple regression with age, sex, disease status and carriers status as predictor variables and log-transformed LTB4 quantities as the response.

[b]The correlation between disease status and LTB4 has been reported previously[1].

of the cohort has been described previously[1]. In brief, individuals with myocardial infarction were recruited from a registry that includes all individuals with myocardial infarction diagnosed before the age of 75 in Iceland from 1981 to 2002, according to WHO-MONICA criteria for acute myocardial infarction[25]. Neurologists and vascular surgeons confirmed the diagnoses of stroke and peripheral vascular disease, respectively.

The Data Protection Commission and the National Bioethics Committee of Iceland approved the study. Informed consent was obtained from all study participants. Personal identifiers were encrypted with a third-party encryption system[26].

Subjects from Philadelphia. Study participants were enrolled at the University of Pennsylvania Medical Center through the PENN CATH study program, which studies the association of biochemical and genetic factors with coronary artery disease in subjects undergoing cardiac catheterization. In total 3,850 subjects have participated. For our study, we selected from the PENN CATH study 833 individuals (728 European Americans and 105 African Americans) diagnosed with myocardial infarction on the basis of either criteria for acute myocardial infarction (an increase in cardiac enzymes and electrocardiographic changes) or a self-reported history of myocardial infarction. For controls, we selected 557 individuals (430 European Americans and 127 African Americans) who showed no evidence of coronary artery disease (luminal stenosis less than 10%) on coronary angiography. Ethnicity information was self-reported.

The University of Pennsylvania Institutional Review Board approved the study, and all subjects provided written informed consent.

Subjects from Cleveland. Study participants were enrolled at the Cleveland Clinic Heart Center through the Genebank program, which is a registry of data and biological samples obtained from individuals undergoing coronary catheterization. The diagnostic criteria for myocardial infarction were based on at least two of the following: prolonged chest pain, electrocardiogram patterns consistent with acute myocardial infarction or a significant increase in cardiac enzymes. Subjects from the Genebank registry who lacked both significant luminal stenosis (<50% stenosis), as assessed by coronary angiography, and a previous history of coronary artery disease were selected as controls for the current study.

The study group included 680 individuals with myocardial infarction (627 European Americans and 53 African Americans) and 903 controls (792 European Americans and 111 African Americans). Ethnicity information was self-reported.

The study was approved by the Cleveland Clinic Foundation Institutional Review Board on Human Subjects, and all subjects gave written informed consent.

Subjects from Atlanta. Study participants were enrolled at the Emory University Hospital, the Emory Clinic and Grady Memorial Hospitals through the Emory Genebank and Clinical Registry in Neurology (CRIN). The Emory Genebank studies the association of biochemical and genetic factors with coronary artery disease in subjects undergoing cardiac catheterization. So far, 736 subjects have participated. For our study, those subjects who had a self-reported history of myocardial infarction (236 European Americans and 39 African Americans) were selected for the myocardial infarction group. Control subjects (553 European Americans and 149 African Americans) were selected from a group of individuals with nonvascular neurological diseases (mainly Parkinson and Alzheimer diseases) recruited from CRIN, their spouses, unrelated friends and community volunteers. These subjects were matched for age and ethnicity to the population with myocardial infarction population. Controls were excluded if they had a known history of myocardial infarction. All subjects provided written informed consent. Information on ethnicity was self-reported.

Statistical analysis. The haplotype association study was done with the program NEMO[27], which handles missing genotypes and uncertainty with phase through a likelihood procedure using the expectation-maximization algorithm to estimate haplotype frequencies. We emphasize that the likelihood ratio tests used explicitly take the uncertainty of the haplotypes counts into consideration, distinguishing them from a two-step procedure that first estimates haplotype counts and then treats the estimated counts as though they are actual counts. The relative risk of a particular haplotype was calculated by a multiplicative model in which the risk of the two alleles of a haplotype that a person carries multiplies[28,29]. With cohort adjustment, the model used for testing was essentially the Mantel-Haenszel test[2], in which each cohort is allowed to have different control haplotype frequencies, but the relative risk is assumed to be the same across cohorts. We extended the standard Mantel-Haenszel test to take into account the incomplete information on haplotype counts. Our admixture adjustment was similar to that proposed in ref. 12, in which the baseline or control frequencies of haplotypes are assumed to be a function of the admixture fraction and a likelihood ratio test is used. Similar to the Mantel-Haenszel model, however, we assumed here that the relative risk is a constant independent of admixture fraction, whereas it is assumed otherwise in ref. 12. The difference is likely to be small here, as we did the admixture adjustment separately in self-reported African Americans and in self-reported European Americans, and not in a combined group.

We used the program Structure[5] to estimate the genetic ancestry of individuals. Structure infers the allele frequencies of K ancestral populations on the basis of multilocus genotypes from a set of individuals and a user-specified value of K, and it assigns a proportion of ancestry from each of the inferred K populations to each individual. Our data set was analyzed by the admixture model, in which the ancestry prior alpha was allowed to vary among populations. This is an important option when genetic material from the K inferred ancestral populations (in this case the African and European ancestral populations) is not equally represented in the data set. This was clearly the case in our data set, which contained 3,366 self-reported European Americans, 584 self-reported African Americans, 364 Icelanders and 87 Nigerians. We ran Structure several times for each value of K in the range 2 to 5. We used the Icelanders and European Americans to identify the European ancestry component in the African Americans and the Nigerians to identify the African ancestry component. On the basis of these runs, we found evidence to indicate that $K = 3$ provides the best estimates of European ancestry in African Americans.

First, these estimates corresponded closely to independent group estimates based on Long's WLS admixture estimator[6]. Second, the results obtained with $K = 3$ indicated the existence of clearly defined African and European ancestral gene pools and a third gene pool that contributed a small amount (1–2%) to European and African Americans but nothing to Nigerians and Icelanders. An independent Structure analysis that also included Native American and East Asian reference samples indicated that this third component represented Asian ancestry. When $K > 3$, the European component became divided equally among the additional ancestral gene pools, whereas the African and Asian components remained stable in single components. Thus, $K > 3$ did not seem to provide any additional resolution to the data. Notably, the estimates of European ancestry for African American individuals were strongly correlated between different runs of Structure, regardless of the value of K. Thus, the average Spearman's rank correlation between runs was 0.987 and had a minimum of 0.964. The statistical significance of the difference in mean European ancestry between African American individuals with myocardial infarction and controls was evaluated by reference to a null distribution derived from 10,000 randomized data sets.

To genetically evaluate ancestry of the study cohorts from the US, we selected 75 unlinked microsatellite markers (**Supplementary Table 4** online) from about 2,000 microsatellites genotyped in a multiethnic cohort of 35 European

Table 4 Distribution of genetically determined European ancestry in myocardial infarction case-control cohorts

Cohort	Self-reported ethnicity	Disease status	WLS group estimate of European ancestry (s.e.m.)[a]	Distribution of estimated individual European ancestry[b] Mean	s.d.	Median	25th–75th percentile range
Yoruban Nigerians	African	N/A	N/A	0.036	0.024	0.03	0.019–0.043
Iceland	European	N/A	N/A	0.991	0.015	0.994	0.990–0.996
All American	Eur. Am.	Individuals with MI	0.98 (0.0083)	0.965	0.083	0.991	0.977–0.995
All American	Eur. Am.	Controls	0.979 (0.0079)	0.969	0.07	0.992	0.979–0.995
Philadelphia	Eur. Am.	Individuals with MI	0.974 (0.0101)	0.955	0.101	0.99	0.971–0.995
Philadelphia	Eur. Am.	Controls	0.969 (0.009)	0.959	0.09	0.991	0.969–0.995
Cleveland	Eur. Am.	Individuals with MI	0.982 (0.0079)	0.971	0.068	0.991	0.980–0.995
Cleveland	Eur. Am.	Controls	0.981 (0.0081)	0.972	0.06	0.991	0.979–0.995
Atlanta	Eur. Am.	Individuals with MI	0.995 (0.0075)	0.981	0.038	0.991	0.984–0.994
Atlanta	Eur. Am.	Controls	0.982 (0.0092)	0.973	0.066	0.993	0.983–0.995
All American	Afr. Am.	Individuals with MI	0.243 (0.0138)	0.223	0.184	0.178	0.108–0.282
All American	Afr. Am.	Controls	0.213 (0.016)	0.199	0.145	0.174	0.094–0.267
Philadelphia	Afr. Am.	Individuals with MI	0.252 (0.0178)	0.235	0.195	0.188	0.121–0.288
Philadelphia	Afr. Am.	Controls	0.213 (0.0217)	0.186	0.137	0.157	0.082–0.257
Cleveland	Afr. Am.	Individuals with MI	0.232 (0.0222)	0.21	0.174	0.16	0.096–0.282
Cleveland	Afr. Am.	Controls	0.239 (0.0219)	0.223	0.136	0.191	0.127–0.281
Atlanta	Afr. Am.	Individuals with MI	0.226 (0.0246)	0.206	0.166	0.167	0.098–0.283
Atlanta	Afr. Am.	Controls	0.198 (0.0128)	0.193	0.155	0.161	0.086–0.252

[a]Long's WLS measure of admixture[6] was calculated with alleles from the set of 75 microsatellite markers. Frequencies from Icelanders and Nigerians were used to represent the ancestral allele frequencies of the European and African parental gene pools, respectively. In line with ref. 7, only 16 loci with alleles showing large differences in frequency ($\delta \geq 0.5$) between the two parental populations were used. For the African American cohorts, we calculated the WLS admixture statistic using the European American controls from the same city as representatives of the ancestral European gene pool. In each case, the estimate of European ancestry was higher by about 0.01 (data not shown). This is likely to be due to the small fraction of African alleles present in the European Americans and indicates that Icelanders serve as effective representatives of the European component of the European American gene pool. The WLS admixture statistic was also calculated by using alleles of all 75 microsatellite markers, yielding estimates of European ancestry in African Americans that were slightly higher than those reported above (by 0.01–0.02).

[b]Estimates of genetic ancestry were obtained from the Structure software using the parameters and data described in Methods. Note that the output from Structure does not label the ancestral admixture proportions as either 'European', 'African' or 'Asian', but rather as 'inferred cluster 1', 'inferred cluster 2' or 'inferred cluster 3'; however, the distribution of ancestry from these inferred clusters in Icelanders, Nigerians and the American cohorts suggests that they have a relatively straightforward correspondence with the labels 'African', 'European' and 'Asian' ancestry.

Americans, 88 African Americans, 34 Chinese and 29 Mexican Americans[30]. Out of the 2,000 microsatellite markers, the selected set showed the most significant differences among the European Americans, African Americans and Asians, and also had good quality and yield. Thirty-one of these markers have been used for similar purposes elsewhere[16].

Accession codes. GenBank: *LTA4H*, NM_000895.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank the participants who made this study possible; the nurses at the Icelandic Heart Association; personnel at the deCODE core facilities; A. Rosen, T. Kamineni, J. Pareira and the CRIN staff for recruiting subjects; J. Pritchard for discussion on the admixture analyses; and members of the International HapMap Consortium for providing data which were crucial for our analysis. This work was supported by grants from the US National Institutes of Health (NIH) and by the Emory Alzheimer's Disease Center. Research in Atlanta was supported by NIH grant U54 ES012068 and by Emory General Clinic Research Center grant MO1-RR00039. Research in Cleveland was supported by NIH grant P50 HL077107-01.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests (see the *Nature Genetics* website for details).

Published online at http://www.nature.com/naturegenetics
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Helgadottir, A. *et al.* The gene encoding 5-lipoxygenase activating protein confers risk of myocardial infarction and stroke. *Nat. Genet.* **36**, 233–239 (2004).
2. Mantel, N. & Haenszel, W. Statistical aspects of the analysis of data from retrospective studies of disease. *J. Natl. Cancer Inst.* **22**, 719–748 (1959).
3. The International HapMap Consortium. A haplotype map of the human genome. *Nature* **437**, 1299–1320 (2005).
4. Falush, D., Stephens, M. & Pritchard, J.K. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**, 1567–1587 (2003).
5. Pritchard, J.K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).
6. Long, J.C. The genetic structure of admixed populations. *Genetics* **127**, 417–428 (1991).
7. Parra, E.J. *et al.* Estimating African American admixture proportions by use of population-specific alleles. *Am. J. Hum. Genet.* **63**, 1839–1851 (1998).
8. Adams, J. & Ward, R.H. Admixture studies and the detection of selection. *Science* **180**, 1137–1143 (1973).
9. Chakraborty, R., Kamboh, M.I., Nwankwo, M. & Ferrell, R.E. Caucasian genes in American blacks: new data. *Am. J. Hum. Genet.* **50**, 145–155 (1992).
10. McKeigue, P.M., Carpenter, J.R., Parra, E.J. & Shriver, M.D. Estimation of admixture and detection of linkage in admixed populations by a Bayesian approach: application to African-American populations. *Ann. Hum. Genet.* **64**, 171–186 (2000).
11. Reiner, A.P. *et al.* Population structure, admixture, and aging-related phenotypes in African American adults: the Cardiovascular Health Study. *Am. J. Hum. Genet.* **76**, 463–477 (2005).
12. Pritchard, J.K., Stephens, M., Rosenberg, N.A. & Donnelly, P. Association mapping in structured populations. *Am. J. Hum. Genet.* **67**, 170–181 (2000).
13. Bamshad, M., Wooding, S., Salisbury, B.A. & Stephens, J.C. Deconstructing the relationship between genetics and race. *Nat. Rev. Genet.* **5**, 598–609 (2004).
14. Jorde, L.B. & Wooding, S.P. Genetic variation, classification and 'race'. *Nat. Genet.* **36** (Suppl.), S28–S33 (2004).
15. Royal, C.D. & Dunston, G.M. Changing the paradigm from 'race' to human genome variation. *Nat. Genet.* **36**, S5–S7 (2004).
16. Tang, H. *et al.* Genetic structure, self-identified race/ethnicity, and confounding in case-control association studies. *Am. J. Hum. Genet.* **76**, 268–275 (2005).
17. Barnholtz-Sloan, J.S., Chakraborty, R., Sellers, T.A. & Schwartz, A.G. Examining population stratification via individual ancestry estimates versus self-reported race. *Cancer*

Epidemiol. Biomarkers Prev. **14**, 1545–1551 (2005).
18. Hoggart, C.J. *et al.* Control of confounding of genetic associations in stratified populations. *Am. J. Hum. Genet.* **72**, 1492–1504 (2003).
19. Spanbroek, R. *et al.* Expanding expression of the 5-lipoxygenase pathway within the arterial wall during human atherogenesis. *Proc. Natl. Acad. Sci. USA* **100**, 1238–1243 (2003).
20. Aiello, R.J. *et al.* Leukotriene B4 receptor antagonism reduces monocytic foam cells in mice. *Arterioscler. Thromb. Vasc. Biol.* **22**, 443–449 (2002).
21. Mehrabian, M. *et al.* Identification of 5-lipoxygenase as a major gene contributing to atherosclerosis susceptibility in mice. *Circ. Res.* **91**, 120–126 (2002).
22. Samuelsson, B. Leukotrienes: mediators of immediate hypersensitivity reactions and inflammation. *Science* **220**, 568–575 (1983).
23. Nicholls, S.J. & Hazen, S.L. Myeloperoxidase and cardiovascular disease. *Arterioscler. Thromb. Vasc. Biol.* **25**, 1102–1111 (2005).
24. Brennan, M.L. *et al.* Prognostic value of myeloperoxidase in patients with chest pain. *N. Engl. J. Med.* **349**, 1595–1604 (2003).
25. Joint International Society and Federation of Cardiology and World Health Organization Task Force on Standardization of Clinical Nomenclature. Nomenclature and criteria for diagnosis of ischemic heart disease. *Circulation* **59**, 607–609 (1979).
26. Gulcher, J.R., Kristjansson, K., Gudbjartsson, H. & Stefansson, K. Protection of privacy by third-party encryption in genetic research in Iceland. *Eur. J. Hum. Genet.* **8**, 739–742 (2000).
27. Gretarsdottir, S. *et al.* The gene encoding phosphodiesterase 4D confers risk of ischemic stroke. *Nat. Genet.* **35**, 131–138 (2003).
28. Falk, C.T. & Rubinstein, P. Haplotype relative risks: an easy reliable way to construct a proper control sample for risk calculations. *Ann. Hum. Genet.* **51**, 227–233 (1987).
29. Terwilliger, J.D. & Ott, J. A haplotype-based 'haplotype relative risk' approach to detecting allelic associations. *Hum. Hered.* **42**, 337–346 (1992).
30. Smith, M.W. *et al.* A high-density admixture map for disease gene discovery in African Americans. *Am. J. Hum. Genet.* **74**, 1001–1013 (2004).